CLONING AND OVEREXPRESSION OF DEOXYRIBONUCLEASE (NucB) FROM *Bacillus licheniformis* IN *E. coli* EXPRESSION SYSTEM

By

LEONG XIN YUN

Dissertation Submitted in

Partial Fulfilment of the

Requirements for the Degree of

Master of Science (Medical Research)

UNIVERSITI SAINS MALAYSIA

AUGUST 2015

DECLARATION

I hereby declare that I am the sole author of this dissertation entitled 'Cloning and Overexpression of Deoxyribonuclease (NucB) from *Bacillus licheniformis* in *E. coli* Expression System'. I declare that this dissertation is being submitted to Universiti Sains Malaysia (USM) for the purpose of the award of Master of Science (Medical Research). This dissertation is the result of my own research under the supervisiom of Dr. Doblin Anak Sandai and Dr. Muhammad Amir Yunus except as cited in the references. This dissertation contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree.

I authorize Universiti Sains Malaysia (USM) to lend this dissertation to other institutions or individuals for the purpose of scholarly research. I further authorize Universiti Sains Malaysia (USM) to reproduce this dissertation by photocopying or other means, in total or in part, at the request of other institutions or individuals for reference.

LEONG XIN YUN

P-IPM0061/14

ACKNOWLEDGEMENT

First off, thank you to God for giving me the ability to learn, the opportunity to pursue this study and the strength to carry through.

I am grateful to Dr. Doblin Anak Sandai and Dr. Muhammad Amir Yunus for being my supervisors and providing me guidance, advice and support to make improvement all through the study. Thank you to Dr. Ida Shazrina Ismail as well for her constructive comments to help students make progress.

My appreciation also goes to Dr. Citartan Marimuthu for supplying a number of materials needed for the study and Dr. Kumitaa Theva Das for her continuous concern for students' going. To other lab mates and laboratory staffs, special thanks are dedicated for their kindness and time entertaining and assisting me in my hour of need. Besides, thanks to my course mates and friends for sharing their knowledge and giving encouragement cum support from the start.

Also, for my beloved parents and sisters, my utmost gratitude is directed to them for their unconditional love, backing of all forms and care at all times. On the whole, I am appreciative to everyone who directly and indirectly involved in making this happen.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

DNA	deoxyribonucleic acid
RNA	ribonucleic acid
DNase	deoxyribonuclease
β	beta
γ	gamma
%	percentage
bp	base pair
rRNA	ribosomal ribonucleic acid
tRNA	transfer ribonucleic acid
mRNA	messenger ribonucleic acid
eDNA	extracellular deoxyribonucleic acid
EPS	extracellular polymeric substance
kDa	kilodalton
µl	microliter
ng	nanogram
µg	microgram
g	gram
rpm	revolutions per minute
ml	milliliter
°C	degree Celsius
u	micro
LB	lysogeny broth
PCR	polymerase chain reaction
mM	milli molar
µM	micro molar
M	molar
nm	nanometer
dNTP	deoxynucleotide triphosphates
IPTG	isopropyl β-D-1-thiogalactopyranoside

SDS PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
dH_2O	distilled water
TEMED	tetramethylethylenediamine
v/v	volume/volume
w/v	weight/volume
V	voltage
kb	kilobase
HCl	hydrogen chloride
$MgCl_2$	magnesium chloride
NaCl	sodium chloride
NaH_2PO_4	sodium dihydrogen phosphate
Na_2HPO_4	disodium hydrogen phosphate
PBS	phosphate buffered saline
Kan	kanamycin
Amp	ampicillin
His	histidine
BSA	bovine serum albumin

PENGKLONAN DAN EKSPRESI TERLEBIH DEOKSIRIBONUKLEASE (NucB) DARIPADA *Bacillus licheniformis* DALAM SISTEM EKSPRESI *E. coli*

ABSTRAK

Deoksiribonuklease NucB daripada *Bacillus licheniformis* ATCC 14580 dijangka berpotensi menguraikan biofilem mikrob iaitu struktur biologi yang boleh menyebabkan banyak komplikasi buruk dalam konteks industri dan kesihatan kerana ia mempunyai 94% persamaan dengan *B. licheniformis* strain EI- 34-6 yang dipencil dari laut dan telah menunjukkan keupayaan untuk menguraikan biofilem yang dibentuk oleh kedua-dua golongan bakteria Gram-positif dan Gram-negatif dengan cepat. Deoksiribonuklease NucB merupakan nuklease luar sel yang dihasilkan oleh gen NucB yang mempunyai 429 pasangan asas nukleotida dengan saiz 15.332 kDa. Mekanisme tindakannya dalam penguraian biofilem dianggap berkaitan dengan pencernaan satu komponen penting dalam bahan polimer luar sel biofilem iaitu DNA luar sel yang berfungsi mengekalkan integriti struktur biofilem. Pencernaan DNA luar sel menyebabkan ketidakstabilan matriks bahan polimer luar sel biofilem dan seterusnya menguraikan biofilem secara lengkap. Selain itu, NucB daripada *Bacillus* dikatakan mempunyai potensi yang lebih tinggi berbanding dengan nuklease-nuklease lain termasuk DNase I dalam pencernaan DNA luar sel. Tujuan kajian ini dijalankan adalah untuk mengklon dan mengekspres NucB deoksiribonuklease dengan menggunakan plasmid pET dan sistem ekspresi protein *E. coli*. Protein rekombinan akan dilabel dengan tag asid amino 6× histidine di terminal karboksil bagi memudahkan pengesanan dan purifikasi. Konstruk rekombinan pada mulanya dibentuk dan kemudian dimasukkan dalam *E. coli* untuk mengekspres protein dengan menggunakan kedua-dua

induksi konvensional dengan IPTG dan autoinduksi. Perbezaan tahap ekspresi antara sampel induksi dan tanpa induksi pada mulanya dinilai melalui kajian SDS PAGE. Seterusnya, penampilan protein rekombinan disahkan dengan analisis Western blot.

CLONING AND OVEREXPRESSION OF DEOXYRIBONUCLEASE (NucB) FROM *Bacillus licheniformis* IN *E. coli* EXPRESSION SYSTEM

ABSTRACT

NucB deoxyribonuclease of *Bacillus licheniformis* ATCC 14580 is expected to disperse the microbial biofilm which is a problematic biological structure notorious for causing numerous complications in a variety of industrial and healthcare context due to its 94% similarity with that of the marine-isolated *B. licheniformis* strain EI-34-6 in which the latter has demonstrated the ability to rapidly break up the biofilms formed by both Gram-positive and Gram-negative bacteria. It is an extracellular nuclease encoded by the 429 base pairs NucB gene and assuming the size of 15.332 kDa. Its mechanism of action in relation to biofilm dispersal is thought to be the digestion of one important component of the biofilm extracellular polymeric substances (EPS) responsible for the biofilm structural integrity namely the extracellular DNA (eDNA). Digestion of the eDNA causes loosening of the EPS matrix leading to total removal of the biofilm. Besides, the *Bacillus* NucB was claimed to have higher potency compared to other nucleases including DNase I in degrading the eDNA. This study focused to clone and overexpress the NucB deoxyribonuclease using the pET vector and *E. coli* expression system in which the recombinant protein will be tagged with the fusion partner, a 6× polyhistidine at the carboxyl terminal to facilitate detection and purification. The recombinant construct was first formed and subsequently transformed into the *E.coli* expression host for expression using both conventional IPTG induction and autoinduction. The difference in the expression level between the induced and the

uninduced samples was initially observed through SDS PAGE study. The final Western blot analysis showed the presence of the His-tagged target protein.

CHAPTER 1

INTRODUCTION

1.1 Literature review

1.1.1 Nucleases

Nucleases belong to the hydrolases enzymatic group capable of cleaving the phosphodiester bonds between the nucleotide subunits of nucleic acids (Ashok Kumar and Kannan, 2011). They are commonly found in prokaryotes and eukaryotes and are diverse based on their localization, structure, function and mode of action. Nucleases can be broadly categorised into sugar-specific nucleases and sugar non-specific nucleases according to the substrates they hydrolyze. Within sugar specific nucleases which cleave either DNA or RNA, there are deoxyribonucleases and ribonucleases. Deoxyribonucleases further branch into topoisomerases, recombinases, restriction endonucleases and damage-specific nucleases all of which are protein nucleases. For non-protein nucleases, there are ribozymes or RNA splicing enzymes (Mishra, 1995).

Deoxyribonucleases and ribonucleases can be either endonucleases or exonucleases depending on the type of nucleolytic attack. Generally, endonucleases are enzymes that cleave DNA from within the molecule while exonucleases cleave at the ends of the molecule. Deoxyribonucleases are the enzymes that can break phosphodiester linkages of DNA (Sheikh and Rosseini, 2014) while ribonucleases are enzymes that break down RNA. Topoisomerases are enzymes that control the topology which is the winding of DNA at certain steps during DNA replication in prokaryotic and eukaryotic cells (Lodish *et al.*,

2000). Recombinases are nucleases that promote dsDNA breaks repair and allow genetic recombination in the recombination hot-spots vicinity (Jockovich and Myers, 2001). Restriction endonucleases are enzymes that cleave DNA at specific recognition sites (Raghavendra and Rao, 2003) whilst damage-specific nucleases are nucleases that act on damaged DNA to play a central role in DNA repair. Sugar non-specific nucleases hydrolyze both DNA and RNA without showing definite base preferences and they are multifunctional enzymes that are distributed extensively and have functions in nutrition, DNA replication, recombination and repair (Rangarajan and Shankar, 2001). Ribozymes are RNAs with catalytic activity analogous to that of protein enzymes (Talini *et al.*, 2011).

Nucleases are evidenced to play crucial roles in range of cellular and genetic mechanisms. These include the regulation of processes such as DNA replication, repair, recombination and mutagenesis (Baker and Kornberg, 1992). Besides, nucleases are also found to be implicated in the control of gene expression by deciding the nature and turnover of gene transcripts. In addition, gene transposition and rearrangements also require the activity of nucleases (Mishra, 1995). The role in host defense against foreign nucleic acids to protect the integrity of host genomes is as well one of the salient features of nucleases (Hsia *et al.*, 2005).

In both eukaryotes and prokaryotes, the various biological roles of nucleases and their applications have been elucidated to great extent (Mishra, 1995). While the functions of nucleases were extensively understood within a cell, which predominantly are related to DNA replication and repair, RNA maturation, restriction and molecular applications (Marti and Fleck, 2004), information on the diversity and importance of extracellular nucleases in

the natural environment is scarce. However, there were a number of different types of extracellular nucleases produced by assorted groups of microbial communities that have been identified in researches carried out previously (Rajarajan, 2013).

1.1.1.1 Bacterial extracellular DNase

Extracellular enzymes or exoenzymes are enzymes which are totally separated from the cell and found free in the surrounding medium yet the division is marginal between them and the cell wall or membrane-bound enzymes (Priest, 1977). Despite the fact that extracellular DNases are produced by numerous pathogenic bacteria, the advantage of enzymatic activity of extracellular DNases is not well understood (Sumby *et al.,* 2005). It was assumed that DNase secretion allows DNA hydrolysis which in turn provides a growth advantage due to enlargement of the pool of available nucleotides (Fox and Holtman, 1968). Also, extracellular DNase activity was proposed to be involved in the propagation of bacteria. It has been hypothesized to be responsible for liquifying pus to disseminate and spread infecting bacteria (Sherry and Goeller, 1950).

Furthermore, in a current study, extracellular DNases were claimed to play a part in degrading neutrophil extracellular traps which are composed of chromatin and granule proteins that can trap and kill bacteria extracellularly and therefore contribute to evasion of the innate immune response (Brinkmann *et al.*, 2004). The exogenous addition of extracellular DNase was demonstrated to have significantly diminished the bactericidal activity of neutrophils directed against *Shigella flexneri* and *Staphylococcus aureus* (Brinkmann *et al.,* 2004). Another discovery of the distinguished function of extracellular DNases is their ability to break up established biofilms and to prevent de novo formation of

biofilms of competitors (Nijland *et al.*, 2010). The DNases can lead to reduction of biofilm and total bacterial biomass, decreased viability of bacteria and declined tolerance to antibiotics which could cause eradication of other competitive bacterial species (Tetz and Tetz, 2010).

1.1.1.2 NucB DNase of *Bacillus licheniformis* ATCC 14580

1.1.1.2.1 B. licheniformis ATCC 14580

B. licheniformis is a Gram-positive bacterium, an endospore former that can be isolated worldwide from soils and plant materials (Sneath *et al.*, 1986). It was suggested to be widely distributed in the environment in form of a saprophytic organism (Alexander, 1977). *B. licheniformis* is closely related to *B. subtilis* and was considered to belong to the *B. subtilis* group II along with other recognized species whose complete genome sequence has been resolved such as *B. anthracis* (Read *et al*., 2003), *B. cereus* (Rasko *et al*., 2004), *B. thuringiensis*, *B. halodurans* (Takami *et al.*, 2000) and *B. subtilis* (Kunst *et al*., 1997). The genome features of *B. licheniformis* and comparison with other *Bacillus* species is shown in Table 1.1.

Table 1.1: Genome features of *B. licheniformis* and comparison with other *Bacillus* species.

Feature	*B. licheniformis*	*B. subtilis**	*B. halodurans*†	*Oceanobacillus iheyensis*‡	*B. anthracis*§	*B. cereus*¶
Chromosome size (bp)	4,222,336	4,214,630	4,202,352	3,630,528	5,227,293	5,426,909
G+C content (mol%)	46.2	43.5	43.7	35.7	35.4	35.4
Protein coding sequences	4,208	4,106	4,066	3,496	5,508	5,366
Average length (bp)	873	896	879	883	800	835
Percent of coding region	86	87	85	85	84	84
Ribosomal RNA operons	7	10	8	7	11	13
Number of tRNAs	72	86	78	69	95	108
Phage-associated genes	71	268	42	27	62	124
Transposase genes of IS elements	10	0	93	14	18	10

[Source: Rey *et al.*, 2004]

It is an organism that has been studied in detailed second only to *E. coli*. *B. licheniformis* is a facultative anaerobe in contrast to majority of the bacilli which are mostly aerobic and this trait may allow it to tolerate and grow in additional wider ecological niches. There was no report that the organism is pathogenic for either animals or plants. It is widely used for producing exoenzymes in large-scale due to its capability of secreting large quantities of proteins of up to 20-25g/l (Schallmey *et al.*, 2004). *B. licheniformis* and its extracellular products have abundant commercial and agricultural applications. Many industrial enzymes such as proteases, α-amylase, penicillinase, pentosanase, cycloglucosyltransferase, β-mannanase and several pectinolytic enzymes have been manufactured using the species (Erickson, 1976). In addition to various specialty chemicals like citric acid, inosine, inosinic acid and poly-γ-glutamic acid, peptide antibiotics for example bacitracin and proticin are produced by specific strains of *B. licheniformis* (Pienta *et al.*, 1989). Several isolates of *B. licheniformis* can also be employed to tone down fungal pathogens that attack the maize, grasses and vegetable crops (Neyra *et al.*, 1996).

According to (Rey *et al.*, 2004), the genome of *B. licheniformis* strain ATCC 14580 has 4,222,336 base pairs and comprises 4208 predicted protein-coding sequences. The overall average genomic guanosine and cytosine content is 46.2%. Besides, there were no plasmids identified during the genome analysis and agarose gel electrophoresis. The protein-coding sequences were stated to have occupied 87% of the genome. They have lengths ranging from 78bp to 10,767bp, with 873bp on average. The sequences are mainly oriented in the direction of replication on the chromosome in which 74.4% of the genes are on the leading strand while the remaining 25.6% reside on the lagging strand (Rey *et al.*, 2004).

Based on the updated information in NCBI, *B. licheniformis* strain ATCC 14580 has the NCBI reference sequence number of NC_006270.3. Its genome is a circular chromosome with 4222597 base pairs and a total of 4332 genes. Besides, the number of true protein coding sequences is 4173. There are also 66 pseudogenes, 21 rRNAs and 72 tRNAs being reported (NCBI, 2014).

Among the 4,208 predicted protein coding genes according to Rey *et al*. (2014), 3948 genes have high similarity to proteins in the non-redundant annotated protein sequence database, Protein Information Resource. 76% of the 3948 gene models contain Interpro motifs and 69% contain protein motifs found in the PFAM database of protein families. There were also 1106 hypothetical proteins and 212 conserved hypothetical proteins in the *B. licheniformis* genome with hits in the Protein Information Resource database (Rey *et al.*, 2004).

The SignalP server which functions to predict the presence and location of signal peptide cleavage sites in protein sequences from diverse organisms has estimated that 689 of the 4208 gene models in the *B. licheniformis* genome have signal peptides (Nielsen *et al.*, 1997). Signal peptides are short peptides present at the N-terminus of most of the newly synthesized proteins that are directed to proceed to the secretory pathway (Blobel and Dobberstein, 1975). Of the 689 genes, the TransMembrane prediction using Hidden Markov Models server (TMHMM) has predicted that 309 have no transmembrane domain (Krogh *et al.*, 2001) and 134 are hypothetical or conserved hypothetical genes. Meanwhile, for the remaining 175 genes, at least 82 are expected to encode secreted proteins and enzymes after a manual examination. Additionally, *B. licheniformis* ATCC 14580genome was also proposed to encode 27 predicted extracellular proteins that are not found in the *B.*

subtilis. Being a saprophytic organism, *B. licheniformis* was also assumed to encode multitude of secreted enzymes involved in hydrolysis of polysaccharides, proteins, lipids and other nutrients (Rey *et al.*, 2004).

1.1.1.2.2 NucB DNase

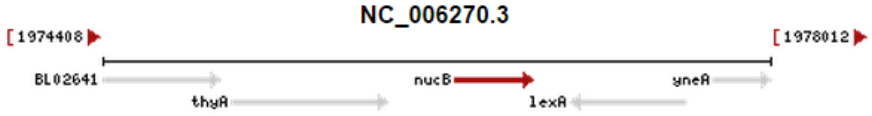
The NucB DNase from *B. licheniformis* ATCC 14580 is an extracellular DNase. The complete gene sequence has a total size of 429 base pairs as shown in Figure 1.1. It is a linear DNA sequence located in the bacterial chromosome from base 1976301 to base 1976729 as shown in Figure 1.2.

```
  1 atgatcaaaa aatgggcggt tcatctgctg ttttccgcat tggtactgct
    tgggctttcg
 61 ggaggcgccg catattctcc tcagcatgcc gaaggtgctg caaggtatga
    cgacatattg
121 tattttccgg catcacgcta tcccgaaacc ggcgctcata tcagcgacgc
    aatcaaagca
181 gggcattcag atgtctgcac gattgaaaga tcgggagcgg ataagcgccg
    ccaggaatca
241 ctgaagggga ttccgactaa gccgggcttt gaccgtgacg aatggccgat
    ggccatgtgt
301 gaagaagggg gcaaaggagc gtctgtcaga tatgtcagct catcggataa
    ccgcggagcc
361 ggctcctggg tcgggaacag gctgagcggt ttcgccgacg ggacgagaat
    tttgtttatc
421 gttcaataa
```

[Source: NCBI, 2013]

Figure 1.1: *B. licheniformis* ATCC 14580 chromosomal NucB gene sequence.

Sequence: NC_006270.3 (1976301..1976729)



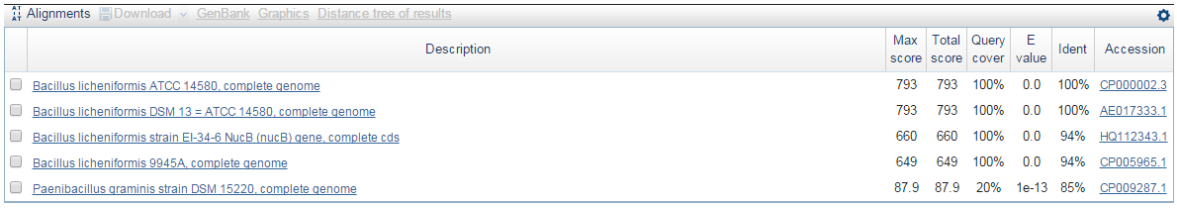
[Source: NCBI, 2014]

Figure 1.2: *B. licheniformis* ATCC 14580 chromosomal NucB gene location.

The nucleotide blast (blastn) performed using the *B. licheniformis* ATCC 14580 chromosomal NucB gene sequence as entry query sequence and nucleotide collection (nr/nt) as the search database produced the BLAST search result as shown in Figure 1.3. The blast output showed that the query sequence has similarity with several sequences.

The NucB gene sequence of *B. licheniformis* ATCC 14580 (Accession no: CP000002.3) and that of *B. licheniformis* DSM 13(Accession no: AE017333.1) have 100% identity. The *B. licheniformis* strain EI-34-6 has only the established NucB gene sequence in the database but not the whole genome sequence. Its NucB gene complete coding sequence (Accession no: HQ112343.1) is identical to that of *B. licheniformis* ATCC 14580 strain by 94%. The same goes to *B. licheniformis* 9945A (Accession no: CP005965.1) in which the identity to the query sequence is 94%. For *Paenibacillus graminis* strain DSM 15220 (Accession no: CP009287.1), only 20% of the query sequence is aligned with which gives rise to 85% identity (NCBI, 2014).

In term of protein, the NucB DNase is composed of 142 amino acids as shown in Figure 1.4 with the GenPept, NCBI accession number of Q65J43 (NCBI, 2006). Based on the information from Uniprot, the protein has the molecular mass of 15.332 kDa and is a member of the deoxyribonuclease NucA/NucB family which has the deoxyribonuclease NucA/NucB domain and acts as a deoxyribonuclease (UniProt, 2004).



Alignments | Download | GenBank | Graphics | Distance tree of results

Description	Max score	Total score	Query cover	E value	Ident	Accession
Bacillus licheniformis ATCC 14580, complete genome	793	793	100%	0.0	100%	CP000002.3
Bacillus licheniformis DSM 13 = ATCC 14580, complete genome	793	793	100%	0.0	100%	AE017333.1
Bacillus licheniformis strain EI-34-6 NucB (nucB) gene, complete cds	660	660	100%	0.0	94%	HQ112343.1
Bacillus licheniformis 9945A, complete genome	649	649	100%	0.0	94%	CP005965.1
Paenibacillus graminis strain DSM 15220, complete genome	87.9	87.9	20%	1e-13	85%	CP009287.1

[Source: NCBI, 2014]

Figure 1.3: Distribution of 5 blast hits on *B. licheniformis* ATCC 14580 chromosomal NucB gene.

1 mikkwavhll fsalvllgls ggaayspqha egaaryddil yfpasrypet gahisdaika

61 ghsdvctier sgadkrrqes lkgiptkpgf drdewpmamc eeggkgasvr yvsssdnrga

121 gswvgnrlsg fadgtrilfi vq

[Source: NCBI, 2006]

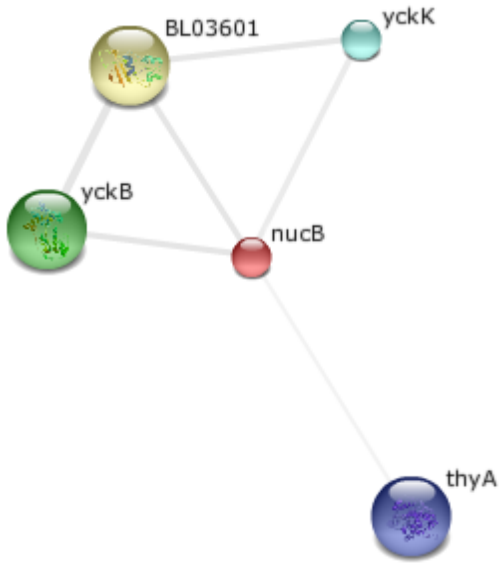
Figure 1.4: *B. licheniformis* ATCC 14580 NucB DNase amino acid sequence.

According to the BioCyc Database Collection, NucB is a sporulation-specific extracellular nuclease (BioCyc, 2014). Besides, based on the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) known and predicted protein-protein interaction database, the NucB DNase is estimated to have four predicted functional partners with the network display illustrated in Figure 1.5. The functional partners are thyA (thymidylate synthase) which has 279 amino acids that provides the sole de novo source of deoxythymidine monophosphate for DNA biosynthesis, YckB protein with 291 amino acids which is predicted to be the ABC transporter substrate-binding protein, YckK protein with 268 amino acids which has transporter activity and finally the hypothetical protein BL03601 which has 151 amino acids (STRING, 2014).

The blastp result shows that NucB DNase has significant alignments with a number of proteins in which the distribution of 100 blast hits on the query sequence has identity ranges from 100% to 55%. The protein with 100% identity to NucB is a sporulation protein from the *Bacillus* without specification of the species and with NCBI accession number of WP_003182220 while the protein having 55% identity to NucB is the sporulation protein from *Paenibacillus* sp. A9 with accession number of WP_026136223.1 (NCBI, 2015).

PSIPRED (Psi-blast based secondary structure prediction) is a highly accurate protein structure prediction server that provides the predicted secondary structure of proteins (McGuffin *et al.*, 2000). The predicted secondary structure of NucB composed of alpha helix, beta strand and random coil along with the legend is shown in Figure 1.6. The iterative threading assembly refinement (I-TASSER) is a bioinformatics approach for automated protein structure and function prediction starting from the amino acid sequence

to structure to function prediction (Roy *et al.*, 2010). The predicted 3-dimensional model of NucB DNase is displayed in Figure 1.7.



[Source: STRING, 2014]

Figure 1.5: *B. licheniformis* ATCC 14580 NucB DNase predicted functional partners network.

[Source: UCL Bioinformatics Group, 2015]

Figure 1.6: Predicted secondary structure of NucB DNase.

[Source: I-TASSER, 2015]

Figure 1.7: Predicted 3-dimensional model of NucB DNase.

1.1.1.2.3 Application of NucB DNase

On the whole, there are two life forms of bacteria during growth and proliferation. The bacteria can either exist as single, independent cells known as the planktonic form or in form of sessile aggregates called biofilm (Bjarnsholt, 2013). Biofilm is a mixture of microbial cells and extracellular polymeric substances made up of polysaccharides, proteins, lipids and extracellular DNA (eDNA) (Flemming *et al.*, 2007). Extensive studies on acute infections caused by pathogenic bacteria have been carried out for more than a decade. However, in recent study, it was found that a new class of chronic infections caused by bacteria growing in the biofilm form contributes to bacterial pathogenesis. High death rate was observed as a result of biofilm infections for example pneumonia in patients suffering from cystic fibrosis, chronic wounds, chronic otitis media and implant- and catheter-associated infections which affect millions of people in the developed world each year (Bjarnsholt, 2013). Microbial biofilms hence are noted to have great importance for public health due to their association with certain infectious diseases and a variety of device-related infections (Donlan, 2002).

The formation of biofilms facilitates unicellular organisms to assume a momentary multicellular lifestyle which is important for survival in adverse environments. Such transition to biofilm structure from the planktonic form is triggered by environmental changes and engrosses multiple regulatory networks that in turn convey signals to initiate change in gene expression that finally mediate spatial and temporal reorganization of the bacterial cell (Monds and O'Toole, 2009). Consequence of such cellular restructuring is the alteration of the expression of surface molecules, nutrient consumption, virulence factors

and the bacteria is conferred with multitude of properties crucial for their survival in hostile conditions (Klebensberger *et al.*, 2009).

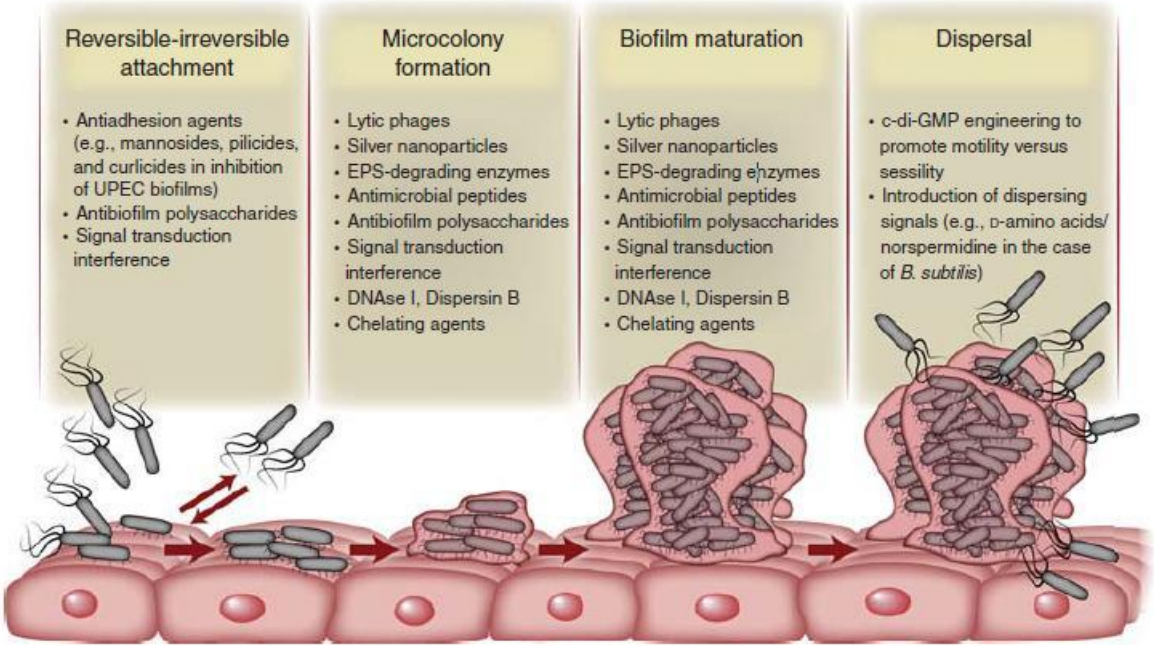
Acute infections were mostly attributable to planktonic bacteria which are usually easily treated with antibiotics even though accurate and fast diagnosis plays essential parts in successful treatment. On the other hand, the infection caused by biofilm bacteria frequently becomes refractory and will develop into a chronic state. The imperative feature of infections caused by bacteria in form of biofilm is the high resistance to antibiotics and majority of conventional antimicrobial agents as well as the exceptional ability to evade the host defences (Bjarnsholt, 2013).

In the current treatment strategy against bacterial infections, use of antibiotics is still the main or favored option. There are two ways conventional antibiotics work against bacteria which are either bacteriostatic that is stopping the bacterial cells from growing or bactericidal that is the killing of the bacterial cells (Kostakioti *et al.*, 2013). Despite antibiotics have been the crucial resolution in eradicating bacterial pathogens over the years, there are multitudes of evidence revealing the damage they render on the host microbiota and thereby facilitating the flourish of opportunistic pathogens which in turn develop higher resistance to antibiotics (Ubeda *et al.*, 2010).

Due to the extracellular matrix of biofilm that poses diffusion limitations, increased transmission of resistance markers within the biofilm community, high metal ion concentration and low pH that inactivate antibiotics and the presence of metabolically inactive persistent cells surviving the treatment, biofilm bacteria become intractable to

antibiotic treatments (Lewis, 2008). Biofilm bacteria thus acquire up to 1000-fold more resistance to antibiotics compared to planktonic cells by the abovementioned features (Høiby *et al.*, 2010). As a consequence, higher effectiveness treatments to overcome biofilm are indispensable. Figure 1.8 shows several strategies developed lately to destroy the bacteria or target certain biofilm developmental stages to prevent biofilm formation (Kostakioti *et al.*, 2013).

NucB deoxyribonuclease from *B. licheniformis* strain EI-34-6 was shown to be able to rapidly disintegrate the biofilms formed by both Gram-positive and Gram-negative bacteria (Nijland *et al.*, 2010). The robust biofilms formed by bacteria involved in chronic rhinosinusitis could also be diminished through NucB treatment in vitro (Shields *et al.*, 2013). NucB deoxyribonuclease has an effective mechanism of action in dispersing preformed biofilms and was suggested to be released as a significant enzyme to degrade the biofilm (Shakir *et al.*, 2012). The nuclease was predicted to digest the eDNA and cause loosening of the EPS matrix which finally leads to total removal of the biofilms. Besides, NucB was claimed to be more superior in degrading eDNA compared to other nucleases including DNase I (Rajarajan *et al.*, 2013). As NucB of *B. licheniformis* ATCC 14580 is as well an extracellular deoxyribonuclease and share 94% similarity with the NucB deoxyribonuclease from *B. licheniformis* strain EI-34-6, it is expected to have the ability to digest the eDNA of biofilm to cause dispersal.



[Source: Kostakioti *et al.*, 2013]

Figure 1.8: Stages in biofilm development and strategies to disrupt biofilm formation at specific stages.

1.1.2 Gene cloning

Gene cloning is the production of multiple copies of a single gene using recombinant DNA technology which is an umbrella term involving the transfer of genetic information from one organism to another (Glick *et al.*, 2010). Recombinant DNA technology has an overall aim of identifying, isolating, manipulating and re-expressing genes from a given host (Lewin, 1990). The practical goals of such cut-and-paste technology include developing a fundamental understanding of the function and regulation of known gene products, identifying new genes whose proteins have been isolated as in reverse genetics, rectifying endogenous genetic defects, expressing foreign genes in disease-susceptible hosts and manufacturing a protein product in large quantities for particular application (Carroll, 1993). The objectives can be met by a variety of methods yet a recombinant DNA experiment usually follows a certain typical steps (Glick *et al.*, 2010).

In order to clone a DNA fragment, the recombinant DNA molecule must be created and copied many times using two essentially different DNAs namely the vector and the insert (Krebs *et al.*, 2013). The DNA of interest is first isolated from donor organism. There are various sources of the targeted DNA depending on the end purpose which can be either DNA fragments representing a specific gene or part of a gene or sequences of the entire genome of an organism isolated in form of genomic DNA, complementary DNA, polymerase chain reaction products or chemically synthesized oligonucleotides (Allison, 2009). Both targeted DNA and small autonomously replicating DNA molecules such as bacterial plasmids are treated with same restriction enzymes to produce compatible DNA ends. The plasmids act as carriers or vectors for the targeted DNA. A range of vector systems apart from plasmids such as bacteriophage lambda, cosmid, bacteriophage P1,

bacterial artificial chromosome, P1-bacteriophage-derived artificial chromosome, yeast artificial chromosome and mammalian artificial chromosome each requiring specific host cells are available depending on the size of insert (Glick *et al.*, 2010). The insert and plasmid vector are mixed and treated with enzyme DNA ligase to ligate them, forming the vector-insert DNA construct or recombinant DNA. The recombinant plasmids are added to competent bacterial cells in which under the right conditions, some of the bacteria will take in the plasmid through a process known as transformation. Bacterial cells are plated and allowed to grow into colonies on antibiotic selection plates. The recombinant plasmid is copied as the bacterial cells replicate. Subsequently, individual transformed cells with single recombinant vector will divide into a colony comprising millions of cells, all carrying the same recombinant vector. The recombinant vector can be extracted and analysis of the cloned DNA fragment can be performed. The recombinant vector with correct cloned DNA can be reintroduced into other host cells to carry out specific functions and manipulations (Griffiths *et al.,* 2000). The target gene can also be expressed through transcription and translation in the host cells to produce targeted proteins when required (Shivanand and Noopur, 2010).

1.1.3 Protein expression

Proteins are extensively used in research, medicine and industry yet the extraction of proteins from their natural sources can be complicated and costly (Ma *et al.*, 2003). As the applications requiring high amounts of good-quality proteins hit the market in increasing trend, efficient recombinant protein production strategies are gaining escalating importance. Despite hundreds of proteins are produced at commercial scale, the production of recombinant proteins remains a challenge in many circumstances. Furthermore, it will be